page 1

PORTAL

USPTO

Search: ⦿ The ACM Digital Library   ⦾ The Guide

feature distribution data mining   [SEARCH]

THE ACM DIGITAL LIBRARY

○← Feedback Report a problem Satisfaction survey

Terms used **feature distribution data mining**   Found **100,425** of **167,655**

Sort results by   [relevance ▼]   ● Save results to a Binder   Try an Advanced Search
Display results   [expanded form ▼]   ? Search Tips   Try this search in The ACM Guide
                                              ☐ Open results in a new window

Results 21 - 40 of 200   Result page: previous 1 **2** 3 4 5 6 7 8 9 10   next
Best 200 shown   Relevance scale ☐ ◰ ◰ ◼ ◼

21 Cluster ensembles --- a knowledge reuse framework for combining multiple partitions ◼
Alexander Strehl, Joydeep Ghosh
March 2003 **The Journal of Machine Learning Research,** Volume 3
Publisher: MIT Press

Full text available: 🗎 pdf(842.50 KB)   Additional Information: full citation, abstract, references, citings, index terms

This paper introduces the problem of combining multiple partitionings of a set of objects into a single consolidated clustering *without* accessing the features or algorithms that determined these partitionings. We first identify several application scenarios for the resultant 'knowledge reuse' framework that we call *cluster ensembles*. The cluster ensemble problem is then formalized as a combinatorial optimization problem in terms of shared mutual information. In addition to a direct ...

Keywords: cluster analysis, clustering, consensus functions, ensemble, knowledge reuse, multi-learner systems, mutual information, partitioning, unsupervised learning

22 Research track: Mining concept-drifting data streams using ensemble classifiers ◼
Haixun Wang, Wei Fan, Philip S. Yu, Jiawei Han
August 2003 **Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining**
Publisher: ACM Press

Full text available: 🗎 pdf(234.13 KB)   Additional Information: full citation, abstract, references, citings, index terms

Recently, mining data streams with concept drifts for actionable insights has become an important and challenging task for a wide range of applications including credit card fraud protection, target marketing, network intrusion detection, etc. Conventional knowledge discovery tools are facing two challenges, the overwhelming volume of the streaming data, and the concept drifts. In this paper, we propose a general framework for mining concept-drifting data streams using weighted ensemble classifi ...

Keywords: classifier, classifier ensemble, concept drift, data streams

23 ◼
Industrial session: new data types and algorithms: SVM in oracle database 10g: removing the barriers to widespread adoption of support vector machines

Boriana L. Milenova, Joseph S. Yarmus, Marcos M. Campos
August 2005 **Proceedings of the 31st international conference on Very large data bases VLDB '05**
Publisher: VLDB Endowment
Full text available: pdf(190.75 KB)   Additional Information: full citation, abstract, references, index terms

Contemporary commercial databases are placing an increased emphasis on analytic capabilities. Data mining technology has become crucial in enabling the analysis of large volumes of data. Modern data mining techniques have been shown to have high accuracy and good generalization to novel data. However, achieving results of good quality often requires high levels of user expertise. Support Vector Machines (SVM) is a powerful state-of-the-art data mining algorithm that can address problems not amen ...

24 Magical thinking in data mining: lessons from CoIL challenge 2000
Charles Elkan
August 2001 **Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining**
Publisher: ACM Press

Full text available: pdf(602.56 KB)   Additional Information: full citation, abstract, references, citings, index terms

CoIL challenge 2000 was a supervised learning contest that attracted 43 entries. The authors of 29 entries later wrote explanations of their work. This paper discusses these reports and reaches three main conclusions. First, naive Bayesian classifiers remain competitive in practice: they were used by both the winning entry and the next best entry. Second, identifying feature interactions correctly is important for maximizing predictive accuracy: this was the difference between the winning classi ...

25 Research sessions: query processing II: Efficient k-NN search on vertically decomposed data
Arjen P. de Vries, Nikos Mamoulis, Niels Nes, Martin Kersten
June 2002 **Proceedings of the 2002 ACM SIGMOD international conference on Management of data**
Publisher: ACM Press
Full text available: pdf(1.26 MB)   Additional Information: full citation, abstract, references, index terms

Applications like multimedia retrieval require efficient support for similarity search on large data collections. Yet, nearest neighbor search is a difficult problem in high dimensional spaces, rendering efficient applications hard to realize: index structures degrade rapidly with increasing dimensionality, while sequential search is not an attractive solution for repositories with millions of objects. This paper approaches the problem from a different angle. A solution is sought in an unconvent ...

26 Differentiating data- and text-mining terminology
Jan H. Kroeze, Machdel C. Matthee, Theo J. D. Bothma
September 2003 **Proceedings of the 2003 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology SAICSIT '03**
Publisher: South African Institute for Computer Scientists and Information Technologists
Full text available: pdf(121.29 KB)   Additional Information: full citation, abstract, references, index terms

When a new discipline emerges it usually takes some time and lots of academic discussion before concepts and terms get standardised. Such a new discipline is text mining. In a groundbreaking paper, <i>Untangling text data mining</i>, Hearst [1999] tackled the problem of clarifying text-mining concepts and terminology. This essay aims to build on Hearst's ideas by pointing out some inconsistencies and suggesting an improved and extended categorisation of data- and text-mining tech ...

Keywords: IR, KDD, TDM, algorithms, database queries, documentation, full-text retrieval, information retrieval, knowledge creation, knowledge discovery, knowledge management, languages, measurement, metadata, text data mining, text mining, text-mining, theory

27 Research track poster: Estimating missed actual positives using independent classifiers
Sandeep Mane, Jaideep Srivastava, San-Yih Hwang
August 2005 **Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining KDD '05**
Publisher: ACM Press
Full text available: pdf(760.77 KB)   Additional Information: full citation, abstract, references, index terms

Data mining is increasingly being applied in environments having very high rate of data generation like network intrusion detection [7], where routers generate about 300,000 -- 500,000 connections every minute. In such rare class data domains, the cost of missing a rare-class instance is much higher than that of other classes. However, the high cost for manual labeling of instances, the high rate at which data is collected as well as real-time response constraints do not always allow one to dete ...

Keywords: capture-recapture method, conditional independence of classifiers given class label, conditional independence of features given class label, conditional mutual information, false negative

28 Special issue on special feature: A divisive information theoretic feature clustering algorithm for text classification
Inderjit S. Dhillon, Subramanyam Mallela, Rahul Kumar
March 2003 **The Journal of Machine Learning Research**, Volume 3
Publisher: MIT Press
Full text available: pdf(171.07 KB)   Additional Information: full citation, abstract, citings, index terms

High dimensionality of text can be a deterrent in applying complex learners such as Support Vector Machines to the task of text classification. Feature clustering is a powerful alternative to feature selection for reducing the dimensionality of text data. In this paper we propose a new information-theoretic divisive algorithm for feature/word clustering and apply it to text classification. Existing techniques for such "distributional clustering" of words are agglomerative in nature and result in ...

29 NASA workshop on issues in the application of data mining to scientific data
Jeanne Behnke, Elaine Dobinson
June 2000 **ACM SIGKDD Explorations Newsletter**, Volume 2 Issue 1
Publisher: ACM Press
Full text available: pdf(1.08 MB)   Additional Information: full citation, index terms

Keywords: NASA, data mining, earth science, statistics

30 Research session: streams: Parameter free bursty events detection in text streams
Gabriel Pui Cheong Fung, Jeffrey Xu Yu, Philip S. Yu, Hongjun Lu
August 2005 **Proceedings of the 31st international conference on Very large data bases VLDB '05**
Publisher: VLDB Endowment

Full text available: pdf(1.03 MB)   Additional Information: full citation, abstract, references, index terms

Text classification is a major data mining task. An advanced text classification technique is known as partially supervised text classification, which can build a text classifier using a small set of positive examples only. This leads to our curiosity whether it is possible to find a set of features that can be used to describe the positive examples. Therefore, users do not even need to specify a set of positive examples. As the first step, in this paper, we formalize it as a new problem, called ...

31 Subspace clustering for high dimensional data: a review

Lance Parsons, Ehtesham Haque, Huan Liu

June 2004 **ACM SIGKDD Explorations Newsletter**, Volume 6 Issue 1

Publisher: ACM Press

Full text available: pdf(539.13 KB)   Additional Information: full citation, abstract, references

Subspace clustering is an extension of traditional clustering that seeks to find clusters in different subspaces within a dataset. Often in high dimensional data, many dimensions are irrelevant and can mask existing clusters in noisy data. Feature selection removes irrelevant and redundant dimensions by analyzing the entire dataset. Subspace clustering algorithms localize the search for relevant dimensions allowing them to find clusters that exist in multiple, possibly overlapping subspaces. The ...

Keywords: clustering survey, high dimensional data, projected clustering, subspace clustering

32 Special issue on special feature: An extensive empirical study of feature selection metrics for text classification

George Forman

March 2003 **The Journal of Machine Learning Research**, Volume 3

Publisher: MIT Press

Full text available: pdf(270.38 KB)   Additional Information: full citation, abstract, citings, index terms

Machine learning for text classification is the cornerstone of document categorization, news filtering, document routing, and personalization. In text domains, effective feature selection is essential to make the learning task efficient and more accurate. This paper presents an empirical comparison of twelve feature selection methods (e.g. Information Gain) evaluated on a benchmark of 229 text classification problem instances that were gathered from Reuters, TREC, OHSUMED, etc. The results are a ...

33 Content-based retrieval for multimedia databases: A unified framework for image database clustering and content-based retrieval

Mei-Ling Shyu, Shu-Ching Chen, Min Chen, Chengcui Zhang

November 2004 **Proceedings of the 2nd ACM international workshop on Multimedia databases**

Publisher: ACM Press

Full text available: pdf(291.68 KB)   Additional Information: full citation, abstract, references, index terms

With the proliferation of image data, the need to search and retrieve images efficiently and accurately from a large image database or a collection of image databases has drastically increased. To address such a demand, a unified framework called <i>Markov Model Mediators</i> (MMMs) is proposed in this paper to facilitate conceptual database clustering and to improve the query processing performance by analyzing the summarized knowledge. The unique characteristics of MMMs are that it ...

Keywords: Markov model mediators (MMMs), content-based image retrieval (CBIR), image database clustering

34 Systems support for scalable data mining

William A. Maniatty, Mohammed J. Zaki

December 2000 **ACM SIGKDD Explorations Newsletter**, Volume 2 Issue 2

Publisher: ACM Press

Full text available: pdf(1.13 MB)   Additional Information: full citation, index terms

Keywords: KDD, data mining, large data sets, parallelism

35 Text classification: Enhanced word clustering for hierarchical text classification

Inderjit S. Dhillon, Subramanyam Mallela, Rahul Kumar

July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**

Publisher: ACM Press

Full text available: pdf(993.07 KB)   Additional Information: full citation, abstract, references, citings, index terms

In this paper we propose a new information-theoretic divisive algorithm for word clustering applied to text classification. In previous work, such "distributional clustering" of features has been found to achieve improvements over feature selection in terms of classification accuracy, especially at lower number of features [2, 28]. However the existing clustering techniques are agglomerative in nature and result in (i) sub-optimal word clusters and (ii) high computational cost. In order to expli ...

36 Multivariate discretization of continuous variables for set mining

Stephen D. Bay

August 2000 **Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining**

Publisher: ACM Press

Full text available: pdf(184.86 KB)   Additional Information: full citation, references, index terms

37 Fast supervised dimensionality reduction algorithm with applications to document categorization & retrieval

George Karypis, Eui-Hong (Sam) Han

November 2000 **Proceedings of the ninth international conference on Information and knowledge management**

Publisher: ACM Press

Full text available: pdf(270.71 KB)   Additional Information: full citation, references, citings, index terms

P⊘RTAL
USPTO

Search: ⊙ The ACM Digital Library   ⌒ The Guide

feature distribution data mining   [SEARCH]

THE ACM DIGITAL LIBRARY

i Feedback Report a problem Satisfaction survey

Terms used **feature distribution data mining**   Found **100,425** of **167,655**

Sort results by   [relevance ▾]   ❧ Save results to a Binder   Try an Advanced Search
Display results   [expanded form ▾]   ? Search Tips   Try this search in The ACM Guide
                                        ⌐ Open results in a new window

Results 21 - 40 of 200   Result page: previous 1 **2** 3 4 5 6 7 8 9 10 next
Best 200 shown   Relevance scale ☐ ⬜ ◼ ◼ ◼

21 Cluster ensembles --- a knowledge reuse framework for combining multiple partitions ◼
Alexander Strehl, Joydeep Ghosh
March 2003 **The Journal of Machine Learning Research**, Volume 3
Publisher: MIT Press

Full text available: 📄 pdf(842.50 KB)   Additional Information: full citation, abstract, references, citings, index terms

> This paper introduces the problem of combining multiple partitionings of a set of objects into a single consolidated clustering *without* accessing the features or algorithms that determined these partitionings. We first identify several application scenarios for the resultant 'knowledge reuse' framework that we call *cluster ensembles*. The cluster ensemble problem is then formalized as a combinatorial optimization problem in terms of shared mutual information. In addition to a direct ...

> **Keywords**: cluster analysis, clustering, consensus functions, ensemble, knowledge reuse, multi-learner systems, mutual information, partitioning, unsupervised learning

22 Research track: Mining concept-drifting data streams using ensemble classifiers ◼
Haixun Wang, Wei Fan, Philip S. Yu, Jiawei Han
August 2003 **Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining**
Publisher: ACM Press

Full text available: 📄 pdf(234.13 KB)   Additional Information: full citation, abstract, references, citings, index terms

> Recently, mining data streams with concept drifts for actionable insights has become an important and challenging task for a wide range of applications including credit card fraud protection, target marketing, network intrusion detection, etc. Conventional knowledge discovery tools are facing two challenges, the overwhelming volume of the streaming data, and the concept drifts. In this paper, we propose a general framework for mining concept-drifting data streams using weighted ensemble classifi ...

> **Keywords**: classifier, classifier ensemble, concept drift, data streams

23 ◼
Industrial session: new data types and algorithms: SVM in oracle database 10g: removing the barriers to widespread adoption of support vector machines

Boriana L. Milenova, Joseph S. Yarmus, Marcos M. Campos
August 2005 **Proceedings of the 31st international conference on Very large data bases VLDB '05**
Publisher: VLDB Endowment
Full text available: 📄 pdf(190.75 KB)   Additional Information: full citation, abstract, references, index terms

Contemporary commercial databases are placing an increased emphasis on analytic capabilities. Data mining technology has become crucial in enabling the analysis of large volumes of data. Modern data mining techniques have been shown to have high accuracy and good generalization to novel data. However, achieving results of good quality often requires high levels of user expertise. Support Vector Machines (SVM) is a powerful state-of-the-art data mining algorithm that can address problems not amen ...

24 Magical thinking in data mining: lessons from CoIL challenge 2000

Charles Elkan
August 2001 **Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining**
Publisher: ACM Press

Full text available: 📄 pdf(602.56 KB)   Additional Information: full citation, abstract, references, citings, index terms

CoIL challenge 2000 was a supervised learning contest that attracted 43 entries. The authors of 29 entries later wrote explanations of their work. This paper discusses these reports and reaches three main conclusions. First, naive Bayesian classifiers remain competitive in practice: they were used by both the winning entry and the next best entry. Second, identifying feature interactions correctly is important for maximizing predictive accuracy: this was the difference between the winning classi ...

25 Research sessions: query processing II: Efficient k-NN search on vertically decomposed data

Arjen P. de Vries, Nikos Mamoulis, Niels Nes, Martin Kersten
June 2002 **Proceedings of the 2002 ACM SIGMOD international conference on Management of data**
Publisher: ACM Press
Full text available: 📄 pdf(1.26 MB)   Additional Information: full citation, abstract, references, index terms

Applications like multimedia retrieval require efficient support for similarity search on large data collections. Yet, nearest neighbor search is a difficult problem in high dimensional spaces, rendering efficient applications hard to realize: index structures degrade rapidly with increasing dimensionality, while sequential search is not an attractive solution for repositories with millions of objects. This paper approaches the problem from a different angle. A solution is sought in an unconvent ...

26 Differentiating data- and text-mining terminology
Jan H. Kroeze, Machdel C. Matthee, Theo J. D. Bothma
September 2003 **Proceedings of the 2003 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology SAICSIT '03**
Publisher: South African Institute for Computer Scientists and Information Technologists
Full text available: 📄 pdf(121.29 KB)   Additional Information: full citation, abstract, references, index terms

When a new discipline emerges it usually takes some time and lots of academic discussion before concepts and terms get standardised. Such a new discipline is text mining. In a groundbreaking paper, <i>Untangling text data mining</i>, Hearst [1999] tackled the problem of clarifying text-mining concepts and terminology. This essay aims to build on Hearst's ideas by pointing out some inconsistencies and suggesting an improved and extended categorisation of data- and text-mining tech ...

Keywords: IR, KDD, TDM, algorithms, database queries, documentation, full-text retrieval, information retrieval, knowledge creation, knowledge discovery, knowledge management, languages, measurement, metadata, text data mining, text mining, text-mining, theory

27 Research track poster: Estimating missed actual positives using independent classifiers
Sandeep Mane, Jaideep Srivastava, San-Yih Hwang
August 2005 **Proceeding of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining KDD '05**
Publisher: ACM Press
Full text available: pdf(760.77 KB)   Additional Information: full citation, abstract, references, index terms

Data mining is increasingly being applied in environments having very high rate of data generation like network intrusion detection [7], where routers generate about 300,000 -- 500,000 connections every minute. In such rare class data domains, the cost of missing a rare-class instance is much higher than that of other classes. However, the high cost for manual labeling of instances, the high rate at which data is collected as well as real-time response constraints do not always allow one to dete ...

Keywords: capture-recapture method, conditional independence of classifiers given class label, conditional independence of features given class label, conditional mutual information, false negative

28 Special issue on special feature: A divisive information theoretic feature clustering algorithm for text classification
Inderjit S. Dhillon, Subramanyam Mallela, Rahul Kumar
March 2003 **The Journal of Machine Learning Research**, Volume 3
Publisher: MIT Press
Full text available: pdf(171.07 KB)   Additional Information: full citation, abstract, citings, index terms

High dimensionality of text can be a deterrent in applying complex learners such as Support Vector Machines to the task of text classification. Feature clustering is a powerful alternative to feature selection for reducing the dimensionality of text data. In this paper we propose a new information-theoretic divisive algorithm for feature/word clustering and apply it to text classification. Existing techniques for such "distributional clustering" of words are agglomerative in nature and result in ...

29 NASA workshop on issues in the application of data mining to scientific data
Jeanne Behnke, Elaine Dobinson
June 2000 **ACM SIGKDD Explorations Newsletter**, Volume 2 Issue 1
Publisher: ACM Press
Full text available: pdf(1.08 MB)   Additional Information: full citation, index terms

Keywords: NASA, data mining, earth science, statistics

30 Research session: streams: Parameter free bursty events detection in text streams
Gabriel Pui Cheong Fung, Jeffrey Xu Yu, Philip S. Yu, Hongjun Lu
August 2005 **Proceedings of the 31st international conference on Very large data bases VLDB '05**
Publisher: VLDB Endowment

Full text available: pdf(1.03 MB)   Additional Information: full citation, abstract, references, index terms

Text classification is a major data mining task. An advanced text classification technique is known as partially supervised text classification, which can build a text classifier using a small set of positive examples only. This leads to our curiosity whether it is possible to find a set of features that can be used to describe the positive examples. Therefore, users do not even need to specify a set of positive examples. As the first step, in this paper, we formalize it as a new problem, called ...

31 Subspace clustering for high dimensional data: a review

Lance Parsons, Ehtesham Haque, Huan Liu

June 2004 **ACM SIGKDD Explorations Newsletter**, Volume 6 Issue 1

Publisher: ACM Press

Full text available: pdf(539.13 KB)   Additional Information: full citation, abstract, references

Subspace clustering is an extension of traditional clustering that seeks to find clusters in different subspaces within a dataset. Often in high dimensional data, many dimensions are irrelevant and can mask existing clusters in noisy data. Feature selection removes irrelevant and redundant dimensions by analyzing the entire dataset. Subspace clustering algorithms localize the search for relevant dimensions allowing them to find clusters that exist in multiple, possibly overlapping subspaces. The ...

Keywords: clustering survey, high dimensional data, projected clustering, subspace clustering

32 Special issue on special feature: An extensive empirical study of feature selection metrics for text classification

George Forman

March 2003 **The Journal of Machine Learning Research**, Volume 3

Publisher: MIT Press

Full text available: pdf(270.38 KB)   Additional Information: full citation, abstract, citings, index terms

Machine learning for text classification is the cornerstone of document categorization, news filtering, document routing, and personalization. In text domains, effective feature selection is essential to make the learning task efficient and more accurate. This paper presents an empirical comparison of twelve feature selection methods (e.g. Information Gain) evaluated on a benchmark of 229 text classification problem instances that were gathered from Reuters, TREC, OHSUMED, etc. The results are a ...

33 Content-based retrieval for multimedia databases: A unified framework for image database clustering and content-based retrieval

Mei-Ling Shyu, Shu-Ching Chen, Min Chen, Chengcui Zhang

November 2004 **Proceedings of the 2nd ACM international workshop on Multimedia databases**

Publisher: ACM Press

Full text available: pdf(291.68 KB)   Additional Information: full citation, abstract, references, index terms

With the proliferation of image data, the need to search and retrieve images efficiently and accurately from a large image database or a collection of image databases has drastically increased. To address such a demand, a unified framework called <i>Markov Model Mediators</i> (MMMs) is proposed in this paper to facilitate conceptual database clustering and to improve the query processing performance by analyzing the summarized knowledge. The unique characteristics of MMMs are that it ...

Keywords: Markov model mediators (MMMs), content-based image retrieval (CBIR), image database clustering

34 Systems support for scalable data mining
William A. Maniatty, Mohammed J. Zaki
December 2000 **ACM SIGKDD Explorations Newsletter**, Volume 2 Issue 2
Publisher: ACM Press
Full text available: pdf(1.13 MB)   Additional Information: full citation, index terms

Keywords: KDD, data mining, large data sets, parallelism

35 Text classification: Enhanced word clustering for hierarchical text classification
Inderjit S. Dhillon, Subramanyam Mallela, Rahul Kumar
July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**
Publisher: ACM Press

Full text available: pdf(993.07 KB)   Additional Information: full citation, abstract, references, citings, index terms

In this paper we propose a new information-theoretic divisive algorithm for word clustering applied to text classification. In previous work, such "distributional clustering" of features has been found to achieve improvements over feature selection in terms of classification accuracy, especially at lower number of features [2, 28]. However the existing clustering techniques are agglomerative in nature and result in (i) sub-optimal word clusters and (ii) high computational cost. In order to expli ...
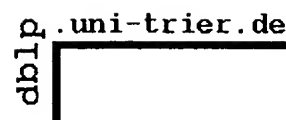
36 Multivariate discretization of continuous variables for set mining
Stephen D. Bay
August 2000 **Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining**
Publisher: ACM Press
Full text available: pdf(184.86 KB)   Additional Information: full citation, references, index terms

37 Fast supervised dimensionality reduction algorithm with applications to document categorization & retrieval
George Karypis, Eui-Hong (Sam) Han
November 2000 **Proceedings of the ninth international conference on Information and knowledge management**
Publisher: ACM Press
Full text available: pdf(270.71 KB)   Additional Information: full citation, references, citings, index terms

Anthology

Very
Large
Data
Bases

dblp .uni-trier.de

21. <u>VLDB</u> 1995: Zurich, Switzerland

<u>Umeshwar Dayal</u>, <u>Peter M. D. Gray</u>, <u>Shojiro Nishio</u>: VLDB'95, Proceedings of 21th International
Conference on Very Large Data Bases, September 11-15, 1995, Zurich, Switzerland. Morgan
Kaufmann, ISBN 1-55860-379-4

```
@proceedings{DBLP:conf/vldb/95,
  editor     = {Umeshwar Dayal and
               Peter M. D. Gray and
               Shojiro Nishio},
  title      = {VLDB'95, Proceedings of 21th International Conference on Very
               Large Data Bases, September 11-15, 1995, Zurich, Switzerland},
  publisher  = {Morgan Kaufmann},
  year       = {1995},
  isbn       = {1-55860-379-4},
  bibsource  = {DBLP, http://dblp.uni-trier.de}
}
```

Very
Large
Data
Bases

Invited Talks

- <u>Henry F. Korth</u>:
  **The Double Life of the Transaction Abstraction: Fundamental Principle and Evolving
  System Concept.** 2-6
  Electronic Edition BibTeX
- <u>Rudolf Bayer</u>:
  Document Management as a Database Problem. 7-10
  Electronic Edition BibTeX
- <u>Stuart E. Madnick</u>:
  **From VLDB to VMLDB (Very MANY Large Data Bases): Dealing with Large-Scale
  Semantic Heterogenity.** 11-16
  Electronic Edition BibTeX

Object Stores

- André Eickler, Carsten Andreas Gerlhof, Donald Kossmann:
  A Performance Evaluation of OID Mapping Techniques. 18-29
  Electronic Edition BibTeX
- Janet L. Wiener, Jeffrey F. Naughton:
  OODB Bulk Loading Revisited: The Partitioned-List Approach. 30-41
  Electronic Edition BibTeX
- Laurent Amsaleg, Michael J. Franklin, Olivier Gruber:
  Efficient Incremental Garbage Collection for Client-Server Object Database Systems. 42-53
  Electronic Edition BibTeX

Distributed Information Retrieval

- David Konopnicki, Oded Shmueli:
  W3QS: A Query System for the World-Wide Web. 54-65
  Electronic Edition BibTeX
- Tak W. Yan, Hector Garcia-Molina:
  Duplicate Removal in Information System Dissemination. 66-77
  Electronic Edition BibTeX
- Luis Gravano, Hector Garcia-Molina:
  Generalizing GlOSS to Vector-Space Databases and Broker Hierarchies. 78-89
  Electronic Edition BibTeX

Disk Management

- Kazuhiko Mogi, Masaru Kitsuregawa:
  Hot Block Clustering for Disk Arrays with Dynamic Striping. 90-99
  Electronic Edition BibTeX
- Kun-Lung Wu, Philip S. Yu, Jen-Yao Chung, James Z. Teng:
  A Performance Study of Workfile Disk Management for Concurrent Mergesorts in a Multiprocessor Database System. 100-109
  Electronic Edition BibTeX
- Ling Tony Chen, Doron Rotem, Sridhar Seshadri:
  Declustering Databases on Heterogeneous Disk Systems. 110-121
  Electronic Edition BibTeX

Concurrency Control

- Azer Bestavros, Spyridon Braoudakis:
  Value-cognizant Speculative Concurrency Control. 122-133
  Electronic Edition BibTeX
- Marcel Kornacker, Douglas Banks:
  High-Concurrency Locking in R-Trees. 134-145
  Electronic Edition BibTeX
- Brajesh Goyal, Jayant R. Haritsa, S. Seshadri, V. Srinivasan:
  Index Concurrency Control in Firm Real-Time Database Systems. 146-157
  Electronic Edition BibTeX

Object Technology

- Peter Buneman, Susan B. Davidson, Kyle Hart, G. Christian Overton, Limsoon Wong:
  A Data Transformation System for Biological Data Sources. 158-169
  Electronic Edition BibTeX
- Fabrizio Ferrandina, Thorsten Meyer, Roberto Zicari, Guy Ferran, Joëlle Madec:
  Schema and Database Evolution in the O2 Object Database System. 170-181
  Electronic Edition BibTeX
- Kazimierz Subieta, Yahiko Kambayashi, Jacek Leszczylowski:
  Procedures in Object-Oriented Query Languages. 182-193
  Electronic Edition BibTeX

Transaction Models

- Marek Rusinkiewicz, Wolfgang Klas, Thomas Tesch, Jürgen Wäsch, Peter Muth:
  Towards a Cooperative Transaction Model - The Cooperative Activity Model. 194-205
  Electronic Edition BibTeX
- Roger S. Barga, Calton Pu:
  A Practical and Modular Implementation of Extended Transaction Models. 206-217
  Electronic Edition BibTeX
- Paul Ammann, Sushil Jajodia, Indrakshi Ray:
  Using Formal Methods to Reason about Semantics-Based Decompositions of Transactions.
  218-227
  Electronic Edition BibTeX

Query Optimization

- Michael Steinbrunn, Klaus Peithner, Guido Moerkotte, Alfons Kemper:
  Bypassing Joins in Disjunctive Queries. 228-238
  Electronic Edition BibTeX
- Waqar Hasan, Rajeev Motwani:
  Coloring Away Communication in Parallel Query Optimization. 239-250
  Electronic Edition BibTeX
- Hongjun Lu, Kian-Lee Tan, Son Dao:
  The Fittest Survives: An Adaptive Approach to Query Optimization. 251-262
  Electronic Edition BibTeX

Multimedia

- Markus Tresch, Neal Palmer, Allen Luniewski:
  Type Classification of Semi-Structured Documents. 263-274
  Electronic Edition BibTeX
- Frank Moser, Achim Kraiss, Wolfgang Klas:
  **L/MRP: A Buffer Management Strategy for Interactive Continuous Data Flows in a
  Multimedia DBMS.** 275-286
  Electronic Edition BibTeX
- Surajit Chaudhuri, Shahram Ghandeharizadeh, Cyrus Shahabi:
  Retrieval of Composite Multimedia Objects. 287-298

Electronic Edition BibTeX

Cost Models

- Alberto Belussi, Christos Faloutsos:
  Estimating the Selectivity of Spatial Queries Using the `Correlation' Fractal Dimension. 299-310
  Electronic Edition BibTeX
- Peter J. Haas, Jeffrey F. Naughton, S. Seshadri, Lynne Stokes:
  Sampling-Based Estimation of the Number of Distinct Values of an Attribute. 311-322
  Electronic Edition BibTeX
- Georges Gardarin, Jean-Robert Gruser, Zhao-Hui Tang:
  A Cost Model for Clustered Object-Oriented Databases. 323-334
  Electronic Edition BibTeX

Query Processing

- Staffan Flodin, Tore Risch:
  Processing Object-Oriented Queries with Invertible Late Bound Functions. 335-344
  Electronic Edition BibTeX
- Weipeng P. Yan, Per-Åke Larson:
  Eager Aggregation and Lazy Aggregation. 345-357
  Electronic Edition BibTeX
- Ashish Gupta, Venky Harinarayan, Dallan Quass:
  Aggregate-Query Processing in Data Warehousing Environments. 358-369
  Electronic Edition BibTeX

Parallel Databases

- Nick Bassiliades, Ioannis P. Vlahavas:
  A Non-Uniform Data Fragmentation Strategy for Parallel Main-Menory Database Systems. 370-381
  Electronic Edition BibTeX
- Manish Mehta, David J. DeWitt:
  Managing Intra-operator Parallelism in Parallel Database Systems. 382-394
  Electronic Edition BibTeX
- Erhard Rahm, Robert Marek:
  Dynamic Multi-Resource Load Balancing in Parallel Database Systems. 395-406
  Electronic Edition BibTeX

Mining Association Rules

- Ramakrishnan Srikant, Rakesh Agrawal:
  Mining Generalized Association Rules. 407-419
  Electronic Edition BibTeX
- Jiawei Han, Yongjian Fu:
  Discovery of Multiple-Level Association Rules from Large Databases. 420-431
  Electronic Edition BibTeX

- Ashok Savasere, Edward Omiecinski, Shamkant B. Navathe:
  An Efficient Algorithm for Mining Association Rules in Large Databases. 432-444
  Electronic Edition BibTeX

Recovery and Availability

- Rainer Gallersdörfer, Matthias Nicola:
  Improving Performance in Replicated Databases through Relaxed Coherency. 445-456
  Electronic Edition BibTeX
- David B. Lomet, Mark R. Tuttle:
  Redo Recovery after System Crashes. 457-468
  Electronic Edition BibTeX
- Svein-Olaf Hvasshovd, Øystein Torbjørnsen, Svein Erik Bratsberg, Per Holager:
  The ClustRa Telecom Database: High Availability, High Throughput, and Real-Time Response. 469-477
  Electronic Edition BibTeX

Data Mining

- Hongjun Lu, Rudy Setiono, Huan Liu:
  NeuroRule: A Connectionist Approach to Data Mining. 478-489
  Electronic Edition BibTeX
- Rakesh Agrawal, King-Ip Lin, Harpreet S. Sawhney, Kyuseok Shim:
  Fast Similarity Search in the Presence of Noise, Scaling, and Translation in Time-Series Databases. 490-501
  Electronic Edition BibTeX
- Rakesh Agrawal, Giuseppe Psaila, Edward L. Wimmers, Mohamed Zaït:
  Querying Shapes of Histories. 502-514
  Electronic Edition BibTeX

Schema Management

- Steven Milliner, Athman Bouguettaya, Mike P. Papazoglou:
  A Scalable Architecture for Autonomous Heterogeneous Database Interactions. 515-526
  Electronic Edition BibTeX
- Eben M. Haber, Yannis E. Ioannidis, Miron Livny:
  OPOSSUM: Desk-Top Schema Management through Customizable Visualization. 527-538
  Electronic Edition BibTeX
- Christian Kalus, Peter Dadam:
  Flexible Relations - Operational Support of Variant Relational Structures. 539-550
  Electronic Edition BibTeX

Access Methods

- Georgios Evangelidis, David B. Lomet, Betty Salzberg:
  The hBP-tree: A Modified hB-tree Supporting Concurrency, Recovery and Node Consolidation. 551-561
  Electronic Edition BibTeX

- Joseph M. Hellerstein, Jeffrey F. Naughton, Avi Pfeffer:
  Generalized Search Trees for Database Systems. 562-573
  Electronic Edition BibTeX
- Sergey Brin:
  Near Neighbor Search in Large Metric Spaces. 574-584
  Electronic Edition BibTeX

Mass Storage

- Sunita Sarawagi:
  Query Processing in Tertiary Memory Databases. 585-596
  Electronic Edition BibTeX
- Luis-Felipe Cabrera, Robert Rees, Wayne Hineman:
  Applying Database Technology in the ADSM Mass Storage System. 597-605
  Electronic Edition BibTeX

Spatial Data

- Erik G. Hoel, Hanan Samet:
  Benchmarking Spatial Join Operations with Spatial Output. 606-618
  Electronic Edition BibTeX
- A. Prasad Sistla, Clement T. Yu, Chengwen Liu, King Liu:
  Similarity based Retrieval of Pictures Using Indices on Spatial Relationships. 619-629
  Electronic Edition BibTeX

Panels

- Raghu Ramakrishnan, Hector Garcia-Molina, Gerhard Rossbach, Abraham Silberschatz, Gio Wiederhold, Jaco Zijlstra:
  Scientific Journals: Extinction or Explosion? (Panel). 631
  Electronic Edition BibTeX
- Andreas Reuter, Stefano Ceri, Jim Gray, Betty Salzberg, Gerhard Weikum:
  Databases and Workflow Management: What is it All About? (Panel). 632
  Electronic Edition BibTeX
- Leonard J. Seligman, Nicholas J. Belkin, Erich J. Neuhold, Michael Stonebraker, Gio Wiederhold:

  Metrics for Accessing Heterogeneous Data: Is There Any Hope? (Panel). 633
  Electronic Edition BibTeX

Application Session 1

- Andreas Meier:
  Providing Database Migration Tools - A Practicioner's Approach. 635-641
  Electronic Edition BibTeX
- Eric Simon, Angelika Kotz Dittrich:
  Promises and Realities of Active Database Systems. 642-653
  Electronic Edition BibTeX

Application Session 2

- Tor Didriksen, César A. Galindo-Legaria, Eirik Dahle:
  Database De-Centralization - A Practical Approach. 654-665
  Electronic Edition BibTeX
- Kazutoshi Sumiya, Kouichi Yasutake, Hirohiko Tanaka, Norio Sanada, Yoshihiko Imai:
  A Product Specification Database for Visual Prototyping. 666-676
  Electronic Edition BibTeX

Vendor Session 1

- David Vaskevitch:
  Very Large Databases: How Large, How Different? 677-685
  Electronic Edition BibTeX

Vendor Session 2

- Yun Wang:
  DB2 Query Parallelism: Staging and Implementation. 686-691
  Electronic Edition BibTeX
- Gilles Fecteau:
  Managing a DB2 Parallel Edition Database. 692-693
  Electronic Edition BibTeX

Vendor Session 3

- Bruce G. Lindsay:
  DB2 Common Server: Technology, Progress, & Directions. 694-695
  Electronic Edition BibTeX
- John Catozzi, Sorana Rabinovici:
  OS Support for VLDBs: Unix Enhancements for the Teradata Data Base. 696-701
  Electronic Edition BibTeX

Vendor Session 4

- Jack A. Orenstein, D. N. Kamber:
  Accessing a Relational Database through an Object-Oriented Database Interface. 702-705
  Electronic Edition BibTeX

Vendor Session 5

- Hannes Spintzik:
  Informix-Online XPS: A Dynamically Scalable RDBMS for Open Parallel Platforms. 706
  Electronic Edition BibTeX
- Gary Hallmark:
  The Oracle Warehouse. 707-709
  Electronic Edition BibTeX

Vendor Session 6

- <u>Harry Leslie</u>, <u>Rohit Jain</u>, <u>Dave Birdsall</u>, <u>Hedieh Yaghmai</u>:
  Efficient Search of Multi-Dimensional B-Trees. 710-719
  Electronic Edition BibTeX
- <u>Paul Brown</u>, <u>Michael Stonebraker</u>:
  BigSur: A System For the Management of Earth Science Data. 720-728
  Electronic Edition BibTeX

<u>ACM SIGMOD Anthology</u> - DBLP: [<u>Home</u> | Search: <u>Author</u>, <u>Title</u> | <u>Conferences</u> | <u>Journals</u>]